Comparative Transcriptomic Analysis Identifies Key Cellulose Synthase Genes (CESA) and Cellulose Synthase-Like Genes (CSL) in Bast Fiber Development Stage of Flax (Linum Usitatissimum L.)

Yuan Guo  
Chinese Academy of Agricultural Sciences

Lan Wen  
The Open University of Hunan

Jikang Chen  
Chinese Academy of Agricultural Sciences

Gen Pan  
Chinese Academy of Agricultural Sciences

Zhimin Wu  
Chinese Academy of Agricultural Sciences

Zheng Li  
Chinese Academy of Agricultural Sciences

Hui Wang  
Central South University of Forestry and Technology

Qingmei Xiao  
Chinese Academy of Agricultural Sciences

Caisheng Qiu  
Chinese Academy of Agricultural Sciences

Songhua Long  
Chinese Academy of Agricultural Sciences

Xinlin Zhao  
Chinese Academy of Agricultural Sciences

Yufu Wang (chinaflax@126.com)  
Chinese Academy of Agricultural Sciences

Research Article

Keywords: Linum usitatissimum L., Cellulose synthase gene, Cellulose synthase-like genes, Bast fiber
Abstract

**Background:** The cellulose synthase gene superfamily, including the cellulose synthase (CESA) and cellulose synthase-like (CSL) gene families, is vital for cell wall construction during plant growth, particularly for fiber development of flax, which is an old and important fiber crop.

**Results:** This study performed a sequencing search of key CESA and CSL genes from several flax stem parts at different fiber development stages by comparing RNA-Seq. Quantitative RT-PCR was used to validate the expression of these genes. This study revealed that CESA4 genes (Lus10008225.g and Lus10008226.g), CESA6 genes (Lus10006161.g and Lus10041063.g), CESA8 genes (Lus10007296.g and Lus10029245.g), CSLD4 gene (Lus10026568.g), CSLE1 (Lus10016625.g) and CSLG genes (Lus10023056.g and Lus10023057.g) were specifically expressed in stem tissue below the snap point where fibers is increased amounts of secondary cell wall deposition. LusCESA4 genes, LusCESA8, genes and LusCSLD4 gene were specifically expressed in fiber development stage during the fast growth period of flax plants. Based on GO and KEGG analyses, it was found that genes involved in pathways of cellulose microfibril organization, galactosyl transferase activity and galactose metabolism were specifically enriched in the stem tissue of the fiber development stage. Other genes involved in cellulose biosynthesis were also analyzed and discussed.

**Conclusion:** The results of this study will provide an important foundation for understanding fiber cell wall biogenesis, particularly the roles of LusCESAs and LusCSLs in the process of fiber development.

**Background**

Flax (*Linum usitatissimum* L.) is an important bast fiber (phloem fiber) crop that has been domesticated and grown by humankind for many years worldwide. Flax bast fibers are the primary phloem fibers, with the base material being cellulose, which determines its physical and chemical properties and process characteristics [1]. The cellulose content of mature fibers may reach to 70% of their weight [2]. As a model of bast fiber crops, flax is also used for studying both bast fiber differentiation and phloem secondary cell wall development due to its small plant size, short life cycle, self-pollination with numerous seeds, small size of diploid genome (approximately 400Mbp), and can be easily grown at high density in the laboratory [3]. In flax, fiber elongation and cell wall deposition are temporally separated, and secondary cell wall deposition begins immediately after fiber elongation has been completed. The transition point in the stem, where fibers have completed elongation and begin secondary cell wall deposition, is called the snap point [4]. During the fast growth period of flax plants, stem regions above the snap point consist of fibers undergoing elongation, and the region surrounding the snap point contains fibers that are in the transition between elongation and cell wall development; in regions below the snap point, fibers had increasing amounts of secondary cell wall deposition. The fast growth period is critical for flax fiber quality because the maximum length of all bast fibers in the mature plant is determined [1,5-7]. Therefore, it is of great significance to study the expression patterns of genes at different developmental stages of
snap point, especially cellulose synthase-related genes, to better understand the development of fiber cell formation and cell wall thickening.

Cellulose is the most important component of plant cell walls for controlling the directional growth of plant cells [8]. It is biosynthesized at the plasma membrane (PM) by the rosette-structured cellulose synthase (CESA) complex (CSC) [9]. Each transmembrane CSC contains 18 or 36 cellulose synthase protein subunits synthesized by β-1,4-glucan chains using UDP-Glc in the cytosol as a substrate [10]. CESAs are a large protein superfamily that includes nine additional subfamilies named cellulose-like (CSL) defined by sequence similarity to CESAs [11,12]. The genes encoding cellulose synthesis were identified in A. xylinum for the first time, and CESA genes have been identified in many other species [13-15]. The Arabidopsis CESA family members are subdivided into six clusters with orthologous genes from other species, reflecting the functional specialization of distinct CESA isoforms [16,17]. Two functional groups among AtCESAs have been distinguished: AtCESA1, AtCESA3, and AtCESA6-like, which contribute to the synthesis of the primary cell wall (PCW) [18,19] and AtCESA4, AtCESA7, and AtCESA8, which are required during the development of secondary cell wall (SCW) [20,21]. The Arabidopsis genome encodes 30 cellulose synthase-like (CSL) genes, which are divided into nine subgroups: CSLA, CSLB, CSLC, CSLD, CSLE, CSLF, CSLG, CSLH and CSLJ, all of which belonging to the glycosyltransferase family 2 (GT2) [22]. CSL genes have been intensively studied due to their implications in the dietary fiber biosynthesis or biomass improvement for bioethanol production [23,24].

High expression levels of CESA genes observed in fast-growing tissues of L. usitatissimum indicate that they are particularly important for this fiber crop [25,26]. Currently, 16 flax cellulose synthase genes have been identified based on their homology to orthologous [26-29]. Two genes, LusCESA7 and LusCESA9, were specific to the stems of fiber flax during the “rapid growth” stage [30]. The expression of secondary wall-type LusCESAs (LusCESA4, LusCESA7, and LusCESA8) was correlated with the expression of two chitinase-like genes (LusCTL1 and LusCTL2), which were highly enriched in xylem [27]. LusCESA4 and LusCESA8 showed higher levels expression in the inner stem tissues than in the outer stem tissues. Both genes were weakly expressed in the flowers, leaves, and roots. LusCESA3 had significantly higher expression in the outer stem tissues [26]. In our previous study, one CESA4 gene, two CESA8 genes, and two CTL2 genes were found to be enriched in the phloem peel library, and one cellulose synthase-like gene (CSLG3) was detected to be enriched in leaves [31]. Forty-five LusCESAs and LusCSLs were found to be involved in the flax process of plant development [32]. LusCSLD3 and LusCSLG3.2/3.3 genes were highly expressed in samples that were in secondary cell wall development at rapid growth and budding stages [33].

To date, several genome-wide gene expression studies using microarrays and high-throughput sequencing technology have been performed on flax bast fibers, and expressed sequence tags (ESTs) of flax bark tissue have been analyzed [34-38]. However, the cell wall is a highly complex structure, and there must be several CESA and CSL genes involved in the fiber cell wall formation process. Therefore, this study aims to perform expression patterns of CESA and CSL genes related to cell wall formation in the process of fiber elongation and cell wall thickening during the rapid growth period of flax by using RNA-
seq analysis and gene expression profiles. Four sequential time and region points are here investigated. The result of this study will provide a fundamental knowledge for the study of secondary growth and fiber formation in economically important plant.

Materials And Methods

Plant material and collection of stem samples
The elite flax variety Zhongya No.1 was used in this study. Flax seeds were planted in pots with a soil layer of 50cm and grown in a greenhouse at a temperature is between 15°C and 22°C. When plant height was approximately 30cm and 50cm, tissue samples of stem apex, which is above the snap point (SP) and mid-section of stem, which is under the snap point, were collected, and all visible leaves were removed. SP is an indicator of fiber developmental transition from intrusive growth to cell wall thickening, and is easily manually detected from the above stem region by the sharp increase of effort required to break the stem [4]. Approximately 3-5 plants per sample were pooled for RNA extraction. All tissue samples collected were frozen immediately in nitrogen and stored at −80°C until further processing. Samples above the SP were named ABP, and samples under the snap point were named UNP (Fig.1). Four biologically independent replicates of ABP30 (ABP30_1, ABP30_2, ABP30_3, and ABP30_4), UNP30 (UNP30_1, UNP30_2, UNP30_3, and UNP30_4), ABP50 (ABP50_1, ABP50_2, ABP50_3, and ABP50_4), UNP50 (UNP50_1, UNP50_2, UNP50_3, and UNP50_4) were sequenced on the Illumina HiSeq platform. The other four biologically independent replicates of ABP30 (ABP30_5, ABP30_6, ABP30_7, and ABP30_8), UNP30 (UNP30_5, UNP30_6, UNP30_7, and UNP30_8), ABP50 (ABP50_5, ABP50_6, ABP50_7, and ABP50_8), UNP50 (UNP50_5, UNP50_6, UNP50_7, and UNP50_8) were used for qRT-PCR analysis.

RNA extraction and sequencing
The collected samples were then ground in liquid nitrogen. Total RNA was extracted from the samples using a plant RNA kit (Omega Bio-Tek, USA), and treated with DNase I (OMEGA BIO-TEK, USA) according to the manufacturer's instructions. RNA quality and purity were determined by denaturing 1% agarose gel electrophoresis and a NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). RNA concentration was measured using the Qubit® RNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). The integrity of the extracted RNA was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

Sixteen DGE libraries (four for ABP30, four for ABP50, four for UNP30, and four for UNP50) were constructed and sequenced using Illumina technology. The experimental summary process was as follows: a total amount of 3 µg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Furthermore, mRNA was purified from total RNA using poly T oligo-attached magnetic beads. Fragmentation was performed using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First-strand cDNA was synthesized
using random hexamer primers and M-MuLV reverse transcriptase (RNase H). Second-strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of the 3' ends of DNA fragments, NEBNext Adaptor with a hairpin loop structure was ligated to prepare for hybridization. To select cDNA fragments of preferentially 250-200 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, USA). Thereafter, 3 µL USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min, followed by 5 min at 95°C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. Finally, PCR products were purified (AMPure XP system), and library quality was assessed using the Agilent Bioanalyzer 2100 system. Clustering of the index-coded samples was performed on a cBot cluster. System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Hiseq platform, and 125bp/150bp paired-end reads were generated.

Reads mapping to the reference genome and quantification of gene expression level

All clean reads were aligned to the flax genome database [39]. The annotation of flax genes was performed using Phytozome (phytozome.jgi.doe.gov; downloaded file: Lusitatissimum_200_v1.0.readme.txt). The index of the reference genome was built using Bowtie v2.2.3, and paired-end clean reads were aligned to the reference genome using TopHat v2.0.12 [40]. TopHat was used as the mapping tool because it can generate a database of splice junctions based on the gene model annotation file and a better mapping result than other non-splice mapping tools [41]. HTSeq v0.6.1 was used to count the read numbers mapped to each gene. Gene expression levels were measured as reads per kilobase of exon model per million mapped reads (RPKM), considering the effect of sequencing depth and gene length for the read count at the same time [42].

Differential expression level analysis

Differential expression analysis of two groups (ABP30 vs. ABP50, ABP30 vs. UNP30, ABP50 vs. UNP50, UNP30 vs. UNP50) was performed using the DESeq R package (1.18.0) [43]. The resulting P-values were adjusted using Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P-value <0.05 found by DESeq were assigned as differentially expressed.

Functional annotation of differentially expressed genes (DEGs)

Gene Ontology (GO) enrichment analysis of differentially expressed genes (DEGs) was performed using the GOseq R package, in which gene length bias was corrected. GO terms with corrected P-values less than 0.05 were considered significantly enriched by differentially expressed genes. KOBAS was used to test the statistical enrichment of differentially expressed genes in the KEGG pathways [44].

Validation of cellulose synthase genes using qRT-PCR analysis
Ten cellulose synthase genes identified by RNA-Seq were assayed using qRT-PCR. Real-time PCR validation was performed on an optical 96-well plate with a PIKOREAL96 real-time PCR system (Thermo, USA). For each sample, the first-strand cDNAs was reverse transcribed from 1.0 μg total RNA treated with DNase I using reverse transcriptase (Cwbiotech, China) according to the manufacturer's instructions. Each sample contained 2.0 μL of a 1/2 dilution of cDNA in water from the reverse-transcribed reaction mentioned above, 1μL of gene-specific sense primer and 1μL anti primer with the concentration of 10 μM, 15μL of UltraSYBR Mixture (CWbiotech, China) in a final volume of 30μL, and then averaged it out to make each reaction is 10μL and in triplicate for each sample. The thermal cycling condition were as follows: 95°C for 10 min, followed by 40 cycles of 95°C for 15s, and 60°C for 30s. From each of the four biologically independent cDNA samples, two independent technical replications were performed and averaged for further calculations. Relative expression levels were determined as previously described [45]. The flax CYC gene was selected as the endogenous control [46]. The primer sequences of the DEGs and CYC genes are listed in Table S1.

**Statistical analysis**

All data are presented as columns with error bars in the figures. All statistical analyses were performed using the statistics software Excel 2010 and SPSS v20.0.

**Results**

**Sequencing mapping and differentially expressed genes**

To characterize the DGE profiles in the ABP and UNP segments of the flax stem, 16 DGE libraries were collected and used for RNA-Seq analysis (Fig.2A). Approximately 47.1–65.0 million raw reads were obtained from these 16 libraries. More than 90% of the reads were mapped to the reference database for the 16 libraries, except UNP30_2. A total of 43,471 genes were detected in all of 16 samples. The RPKM values for each gene in each sample are listed in Table S2. The correlation analysis showed that the correlation between ABP30 and ABP50 was closer than that between ABP30 and UNP30, as well as between ABP30 and UNP50. The correlation between UNP30 and UNP50 was closer than that between ABP30 and UNP30, as well as between ABP30 and UNP50. This indicated that different genes regulate stem growth with different stem regions which above or below the snap point in the same stage. However, the same genes were involved in the growth of the same stem regions at different stem developmental stages. The correlation coefficients of four biologically independent replicates were all above 0.90, except UNP30_1 and UNP30_2, as well as ABP30_1 and ABP30_4, which were 0.899 and 0.829, suggesting that the expression levels of genes in the four replicate libraries were similar and suitable for downstream analysis (Fig.2B).

To identify genes that were differentially expressed in the four tissues of flax, pairs of DGE profiles of the sixteen libraries were compared to analyze gene expression variations. Genes that showed significant differences in expression were identified between the pairwise comparisons of the ABP30 vs. ABP50, ABP30 vs. UNP30, ABP50 vs. UNP50, and UNP30 vs. UNP50 (Fig.2C and Fig.2D). A total of 8639 DEGs
were detected across all four sample types. Among of them, 92% of DEGs were detected in ABP30 vs. UNP30 and UNP30 vs. UNP50. For ABP30 vs. UNP30, there were 2003 up-regulated genes and 1698 down-regulated DEGs. For UNP30 vs. UNP50, there were 2791 up-regulated genes and 1631 down-regulated DEGs. The data suggested that UNP30 had great differences in gene regulation with both ABP30 and UNP50, indicating significant differences between genes that regulate stem regions below the snap point in different stem development stages, including fiber development stage and fiber maturity stages. In the fiber development stage, there are also significant differences between genes that regulate stem regions above the snap point and genes that regulate regions below the snap point. Only 18 up-regulated genes and 12 down-regulated DEGs were detected in ABP30 vs. ABP50, indicating that there are much fewer gene regulating differences than regions below the snap point in different stem development stages. Only 230 up-regulated genes and 256 down-regulated DEGs were detected in ABP50 vs. UNP50, indicating that there are much fewer differences between genes that regulate stem regions above the snap point and genes that regulate regions below the snap point in the fiber mature stage than in fiber development stage. All of detected DEGs with their functional annotations are listed in Table S3.

Most abundant genes in fiber development and maturing stage

For pairwise comparisons of the ABP30 and UNP30, the most abundant transcripts in the stem tissue above the snap point when plant at 30 cm were those of the genes encoding rhamnogalacturonate lyase (Lus10011565.g), 1-aminocyclopropane-1-carboxylate synthase 8 (Lus10002795.g), fasciclin-like arabinogalactan protein (Lus10002984.g, Lus10036112.g, and Lus10036113.g), leucine-rich repeat extensin-like protein 4 (Lus10033672.g), peroxidase (Lus10020994.g and Lus10009932.g), CASP-like protein 1D1 (Lus10012808.g), protein SRG1 (Lus10015252.g), and MLP-like protein 328 (Lus10008932.g). For pairwise comparisons of the UNP30 vs. UNP50, the most abundant transcripts in stem tissue above the snap point when plants at 50 cm (fiber development stage) were those of the genes encoding probable protein kinase At2g41970 (Lus10000773.g), RPW8-like protein 3 (Lus10000836.g), 1-aminocyclopropane-1-carboxylate oxidase 5 (Lus10000857.g), subtilisin-like protease SBT1.6 (Lus10002397.g), cytochrome P450 82C4 (Lus10003148.g), and MLP-like protein 423 (Lus10003451.g). Some genes were highly expressed in the fiber development stage compared to the fiber mature stage but with unknown proteins, such as Lus10000527.g, Lus10000834.g, Lus10000835.g, Lus10002396.g, and Lus10008019.g (Table 1).

For pairwise comparisons of the ABP50 and UNP50, the most abundant transcripts in stem tissue above the snap point when plants were 50 cm were those of the genes encoding MLP-like protein 328 (Lus10008930.g), ethylene-responsive transcription factor ERF022 (Lus10002953.g), cytokinin hydroxylase (Lus10033555.g), high-affinity nitrate transporter 2.4 (Lus10026527.g), cytochrome P450 83B1 (Lus10034505.g), subtilisin-like protease SBT1.7 (Lus10006307.g), dehydration-responsive element-binding protein 1D (Lus10027412.g), and 7-deoxyloganetin glucosyltransferase (Lus10032218.g). Two genes, Lus10015036.g and Lus10041692.g were highly expressed in stem tissue below the snap point with unknown proteins. For pairwise comparisons of the UNP30 vs. UNP50, the most abundant transcripts in stem tissue above the snap point when plants at 50 cm (fiber maturing
stage) were those of the genes encoding 36.4 kDa proline-rich protein (Lus10010482.g), transmembrane protein 205 (Lus10000723.g), auxin-responsive protein SAUR21 (Lus10008992.g), transmembrane protein 205 (Lus10021473.g), BTB/POZ and TAZ domain-containing protein 1 (Lus10033038.g), anthranilate N-methyltransferase (Lus10002668.g), Syntaxin-71 (Lus10007411.g) and short-chain dehydrogenase TIC 32, chloroplastic (Lus10025321.g). Two genes, Lus10015276.g and Lus10033722.g were highly expressed in the fiber mature stage compared to the fiber development stage but with unknown proteins (Table 2).

**Gene ontology and metabolic pathway enrichment analysis of differentially expressed genes**

Using GO analysis, this study classified differentially expressed genes based on enriched GO terms in the biological process, cellular component, and molecular function ontologies. The top 30 significantly enriched categories for the DEGs are listed in Fig. 3. For the stem tissue above the snap point vs. other tissues, cellular components enriched fewer DGEs then the other two ontologies. At the fiber development stage, most of the cellular component genes were enriched above the snap point (ABP30 vs. UNP30). However, at the fiber maturation stage, genes which involved in membrane part, integral component of membrane, intrinsic component of membrane, and intracellular are significantly different between stem tissues above and under the snap point (ABP50 vs. UNP50). For different fiber development stages, most DGEs were enriched in metabolic process, primary metabolic process, catalytic activity, and organic substance metabolic process, and fiber development stage enriched more DEGs than fiber maturing stage for each ontology (UNP30 vs. UNP50). Furthermore, it was found that some cellulose synthesis GO terms are specific in stem tissue of fiber development stage (UNP30), including cellulose microfibril organization (GO:0010215: Lus10034670.g and Lus10017863.g), galactosyl transferase activity (GO:0008378: Lus10023955.g, Lus10021538.g, Lus10039657.g and Lus10027185.g) (Table S4).

Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed to identify major or specific metabolic pathways and their underlying genes in the fiber development stage (Fig. 4). The term “Biosynthesis of secondary metabolites” was significantly enriched in all of the stem tissue above the snap point vs. other tissues, including over 500 DEGs underlying ABP30 vs. UNP30, UNP30 vs. UNP50, and ABP50 vs. UNP50. In particular, the pathways of “Phosphatidylinositol signaling system,” “Zeatin biosynthesis,” “Glycosaminoglycan degradation,” and “Ubiquitin mediated proteolysis“ were significantly enriched in the stem tissue above the snap point. Galactose metabolism pathway is specifically enriched in stem tissue above the snap point in fiber development stage, which include 8 genes (Lus10003875.g, Lus10001822.g, Lus10043044.g, Lus10002740.g, Lus10007370.g, Lus10001529.g, Lus10034218.g, Lus10020788.g) (Table S5), suggesting genes encoding β-galactosidase participated in the early fiber synthesis.

**Differentially expressed genes involved in the cellulose biosynthesis**

Cellulose is synthesized by large multi-meric cellulose synthase (CesA) complexes, tracking along cortical microtubules at the plasma membrane (Fig. 5A). Recent studies have identified the Korrigan (KOR), sucrose synthase (SuSy), and COBRA-like proteins are indirectly involved in cellulose biosynthesis.
According to GO function, this study determined the DEGs involved in the pathway of cellulose synthase activity, cellulose biosynthetic process, cellulose metabolic process, cellulose binding, and cellulose microfibril organization. Most of these genes were highly expressed at the fiber development stage (UBP30), comparing with those at the fiber maturing stage (UBP50) or stem tissues above the snap (ABP30 and ABP50), except Lus10034665.g and Lus10020414.g, which encoding Alpha-1,4-glucan-protein and Phosphoglucosamine mutase (Fig. 5B).

**Analysis of CESA and CSL genes in fiber cell wall formation stage**

Among the 8639 up-regulated and down-regulated DEGs detected across all four types of samples, 15 genes encoding cellulose synthase proteins or cellulose synthase-like proteins were found. These genes were found in pairwise comparisons of ABP30 vs. UNP30 and UNP30 vs. UNP50. Pairwise comparisons of ABP30 vs. ABP50 and ABP50 vs. UNP50 did not reveal genes related to cellulose synthase. In pairwise comparison of ABP30 vs. UNP30, one probable cellulose synthase A catalytic subunit 3 gene (Lus10022449.g) was found to be expressed at significantly higher levels in stem tissue above the snap point than that of below the snap point. In contrast, five cellulose synthase genes, including two CESA4 genes (Lus10008225.g and Lus10008226.g), one CSLD4 gene (Lus10026568.g), and two CESA8 genes (Lus10007296.g and Lus10029245.g), were found to be expressed at significantly higher levels in stem tissue below the snap point than in those above the snap point. In pairwise comparison of UNP30 vs. UNP50, seven cellulose synthase genes, including two CESA8 genes (Lus10007296.g and Lus10029245.g), one CESA4 gene (Lus10008226.g), one CSLE1 gene (Lus10016625.g), two CSLG3 genes (Lus10023056.g and Lus10023057.g), and one CSLD4 gene (Lus10026568.g) were found to be expressed significantly higher in stem tissue with fiber development stage than in the fiber mature stage. In contrast, two CESA6 genes (Lus10006161.g and Lus10041063.g) were found to be expressed at significantly higher levels in stem tissue at the fiber mature stage than at the fiber development stage. To confirm the results of expression profiling by RNA-Seq analysis, ten transcripts of cellulose synthase genes were used as targets for quantitative real-time RT-PCR analysis, including two CESA4 genes (Lus10008225.g and Lus10008226.g), two CESA6 genes (Lus10006161.g and Lus10041063.g), two CESA8 genes (Lus10007296.g and Lus10029245.g), two CSLG genes (Lus10023056.g and Lus10023057.g), CSLE1 (Lus10016625.g), and CSLD4 (Lus10026568.g). The results showed that the expression patterns of two CESA4 genes (Lus10008225.g and Lus10008226.g), two CESA8 genes (Lus10007296.g and Lus10029245.g), CSLE1 (Lus10016625.g), and CSLD4 (Lus10026568.g) were determined by qRT–PCR fitted in well with that of RNA-seq analysis (Fig. 6A). Furthermore, these six genes were highly expressed in stem tissue at the fiber development stage than in the other three samples, suggesting that these genes were enriched during the fiber cell wall formation stage (Fig. 6B).

**Discussion**

The CESA genes
Cellulose synthesis is one of the major ongoing processes in flax phloem fibers during the cell wall formation [36]. It is performed by a multi-enzyme complex containing several types of highly homologous and distinct catalytic subunits, such as CESA glycosyltransferases [47]. In our previous study, one CESA4 (Lus10008226.g) gene and CESA 8 (Lus10007296.g and Lus10029245.g) genes were detected to be enriched in the phloem peel library, revealing that CESA4 and CESA8, which belong to secondary wall-type LusCESAs, are important for phloem fiber development [31]. In the present study, six cellulose synthesis machinery genes were detected in pairwise comparisons of analyzed samples, including one probable cellulose synthase A catalytic subunit 3 (PCESA3) (Lus10022449.g), two CESA4 genes (Lus10008225.g and Lus10008226.g), two CESA6 genes (Lus10006161.g and Lus10041063.g), and two CESA8 genes (Lus10007296.g and Lus10029245.g). PCESA3 was specifically expressed in stem tissue above the snap point. CESA4 genes (Lus10008225.g and Lus10008226.g) and CESA8 genes (Lus10007296.g and Lus10029245.g) were specifically expressed in stem tissue below the snap point when plants were 30cm. CESA6 genes (Lus10006161.g and Lus10041063.g) were specifically expressed in stem tissue below the snap point when plants were 50cm. Similar to our findings, CESA4 and CESA8 were more enriched in the basal region than in the apical region of the flax stem [35]. Despite the very intensive cellulose synthesis in fibers, Gorshkov et al., (2017) reported that none of the CESA genes were significantly up-regulated in fibers relative to their levels in other samples or present among the most highly expressed genes [36]. In a study by Gorshkova et al., (2018), CESA3 (Lus10007538, Lus10012198) and CESA6 (Lus10041063), which belong to primary wall-type LusCESAs, were moderately up-regulating in intrusively growing fibers above the snap point [37]. Construction of co-expression networks for LusCESAs revealed that both primary and secondary cell wall-related CESAs were involved in the joint co-expression group in fibers depositing tertiary cell walls. The set of secondary wall-type LusCESAs (LusCESA4, LusCESA7, and LusCESA8) is always expressed in cells with thick cell walls [48]. In summary, all of the previous studies and our findings suggest that CESA4 genes (Lus10008225.g and Lus10008226.g), CESA8 genes (Lus10007296.g and Lus10029245.g), and CESA6 genes (Lus10006161.g and Lus10041063.g) crucially participated in the regulation of cell wall deposition stage of fibers and should be further studying using gene-editing strategies.

The CSL genes

At present, 30 CSL genes have been subdivided into nine subfamilies: CSLA to CSLJ [47,49,50]. The CSL genes are likely to be responsible for the biosynthesis of the hemicellulose polymers [51]. Among them, CSLA synthesizes the backbone of mannan or glucomannan [52,53]. CSLC catalyzes the synthesis of xyloglucan, which is a major hemicellulose in the primary walls of many land plants [54]. CSLE is only found in angiosperms, and its function is poorly understood until now [51] (Yin et al., 2014). CSLD might act as a mannan synthases, however, the exact polymer products of CLSD are still under debate [51,55]. CSLF and CSLH, which are found only in grasses, are involved in the synthesis of mixed linkage glucans [56,57]. The functions of other CSLs currently remain unknown [57,58]. Twenty-five LusCSL genes have been identified according to the flax and Arabidopsis genomes, including one LusCSLB, 18 LusCSLDs, one LusCSLE, and five LusCSLGs [33]. LusCSLD3 and LusCSLG3.2/3.3 (Lus10023057.g) were strongly expressed in the stem sample below the snap point in seedlings, rapid growth, and budding stages.
According to RNA-Seq data, the LusCSLG genes are expressed in the phloem fibers and their expression is practically absent in the xylem [36,59]. In the present study, two CSLG3 genes (Lus10023056.g and Lus10023057.g) were found to be expressed at significantly higher levels in stem tissue at the fiber development stage than in the fiber mature stage. This result further verified their participation in the flax fiber formation processing. This study hypothesized that CSLG3 encodes a fiber-specific glycosyltransferase involved in cellulose synthases with secondary cell wall-related CESAs, especially CESA4 and CESA8. Furthermore, it was found that LusCSLD4 gene (Lus10026568.g) was specifically expressed in stem tissue below the snap point during the fiber development stage, and its expression was significantly higher both than that of stem tissue with mature fiber stage and stem tissue above the snap point. In evolutionary analysis, CSLD and CESA have a closer evolutionary relationship than other family members [32], indicating that it possibly has a similar effect on the synthesis of mixed-linkage glucan. Therefore, it is reasonable to assume that LusCSLD4 (Lus10026568.g) plays an important role in the early phloem fiber development stage. Aside from LusCSLD, one CSLE1 gene (Lus10016625.g) was also found expressed at significantly higher levels in stem tissue at the fiber development stage than in the fiber mature stage. In a study by Galinousky et al., (2020), LusCSLE1 had the same expression as LusCESA1 and LusCESA6, which encode cellulose synthases responsible for primary cell wall development [33]. It is well known that the primary cell wall from fiber cells should allow the intensive cell wall extension process, which is very important for fiber elongation at the intrusive growth stage [60]. Based on these, the authors can presume that LusCSLE1 participates in the development of the secondary cell wall deposition stage, however, its functions require further research.

Other genes involved in cellulose biosynthesis

Apart from CESA and CSL genes, a number of gene families have been characterized to be involved in cell elongation as well as in the different phases of cell wall deposition and thickening in flax by phylogenetic and expression studies, particularly in fiber cells: β-galactosidase (GH35) [61,62]; UDP-glycosyl transferase[63]; rhamnogalacturonan lyases[64]; pectinesterase (PME and PMEI) [65]; chitinase-like proteins [27]; tubulin and actin for the cytoskeleton [66,67]. In our previous study, it was found that rhamnogalacturonate lyase (RhiE), COBRA-like protein 4 (COBL4), and endo-β-1,4-glucanase (KOR) may play important roles in the secondary cell wall deposition [31]. In this study, COBL4 genes (Lus10017863.g, Lus10031972.g, Lus10035131.g, and Lus10034670.g) and KOR (Lus10012162.g) were moderately up-regulated in stem tissue below the snap point at the fiber development stage (Fig. 5). Gorshkov et al., (2017) found that one COBL4 gene (Lus10035131.g) was significantly up-regulated in phloem fibers isolated from stem peels below the snap point [36]. The COBL4 gene encodes a putative glycosyl-phosphatidylinositol-anchored protein that regulates cellulose deposition and oriented cell expansion in the plant cell wall, and has a carbohydrate-binding, which modules cellulose assembly and cellulose microfibril crystallinity [68]. Based on the above research results, this study believes that COBL4 gene plays an important role in the early stage of secondary cell wall deposition and should draw our attention. The gene encoding rhamnogalacturonate lyase (Lus10004281.g) was highly expressed in stem tissue below the snap point when the plant was 30 cm. These results are consistent with those of Gorshkov et al., (2017) and Gorshkova et al., (2018), in which rhamnogalacturonan lyase protein family genes were
highly expressed during cellulose deposition rather than intrusively growing fibers, which are mainly above the snap point [36,37]. In addition, some genes are probably involved in β-1,4-galactan and β-(1→4)-Gal side-chains to the rhamnogalacturonan lyase backbone, such as GALACTAN SYNTHASE 1 (GALS1). A speculative GALS1 gene (Lus10038387) was highly expressed in stem tissue below the snap point, especially when the plant was 30 cm, and its expression became weak when the plant was 50 cm (Fig. 5). GALS1 specifically forms β-1,4-galactosyl linkages and can add successive β-1,4-galactosyl residues to the acceptor [69]. In a previous study, Lus10038387 showed the most evident pattern of specific transcript enrichment in the phloem fiber sample [36]. Our study also showed that Lus10038387 plays an important role in the early stage of fiber development stage.

**Conclusion**

In this study, the transcriptome profiles of flax stem regions above SP and under SP at two stem development stages, specifically the rapid growth stage and the maturing stage, were investigated to reveal the expression patterns of CESA and CSL genes related to cell wall formation during the process of fiber elongation and cell wall thickening during the rapid growth period of flax. The results were verified using qRT-PCR test and showed that CESA4 (Lus10008225.g and Lus10008226.g), CESA8 genes (Lus10007296.g and Lus10029245.g), CSLD4 (Lus10026568.g), CSLE1 (Lus10016625.g), and CSLG3 (Lus10023056.g and Lus10023057.g) were expressed in stem tissue below SP containing fibers with increasing amounts of secondary cell wall deposition. CESA6 genes (Lus10006161.g and Lus10041063.g) were specifically expressed in stem tissue below SP when fibers were maturing. CESA4 genes, CESA8 genes, and LusCSLD4 gene were specifically expressed in fiber development stage during the fast growth period of flax plants. Expression of other genes that are involved in cellulose biosynthesis, such as COB4 genes (Lus10017863.g, Lus10031972.g, Lus10035131.g, Lus10034670.g), KOR (Lus10012162.g), and GOLS1 (Lus10038387), were also highly and specifically expressed in the stem below the SP at the fiber cell wall thickening stage. This study provides comprehensive information on the CESA and CSL genes and should aid in the understanding of the molecular basis of agronomically important traits of flax, such as fiber quality and other physiological processes.

**Abbreviations**

CESA: cellulose Synthase; CSL: cellulose synthase-like; qRT-PCR: quantitative real-time PCR; SP: the snap point; PM: plasma membrane; CSC: cellulose synthase complex; ESTs: expressed sequence tags; PCW: primary cell wall; SCW: secondary cell wall; GT2: glycosyltransferase family 2; CTL: chitinase-Like; DGE: digital gene expression; DEGs: differentially expressed genes; RPKM: reads per kilobase of exon model per million mapped reads; COBRA: COBRA-like protein; RhiE: rhamnogalacturonate lyase; KOR: endo-β-1,4-glucanase; GALS1: Galactan Synthase 1.

**Declarations**

**Ethics approval and consent to participate**
This study has not directly involved humans or animals. No specific permission was required for use of these materials for experimental purposes. The seedlings were grown in the experimental field of Institute of Bast Fiber Crops, Changsha, China as per standard practices, and samples were harvested at the required time. We comply with the Convention on the Trade in Endangered Species of Wild Fauna and Flora.

Consent for publication

Not applicable.

Availability of data and materials

The datasets for supporting the conclusions of this article are listed in the article and its additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by National Natural Science Foundation of China (grant no. 31771852), China Agriculture Research System Fund (grant no. CARS-16-E13) and the Agricultural Science and Technology Innovation Program in Chinese Academy of Agricultural Sciences (grant no. CAAS-ASTIP-IBFC06). Funding bodies were not involved in the design of the study, collection of data, analysis, and interpretation of data or in writing the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in the article with its supplementary material. We have deposited our transcriptome data in Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra/), the accession number for our submission is: PRJNA723911.

Authors’ contributions

Y.G. performed the experiments and writing-original draft; L. W., J.C., G.P., and Z.W. participated in the sample collection and RNA extraction; Z. L., H. W., Q. X., and C. Q. conducted bioinformatic analysis. S. L. and X. Z. analyzed most of the data; Y. W. provided the materials and financial supporting. All the authors reviewed the draft and approved the final manuscript.

Acknowledgments

We thank all the staff in our lab for providing useful suggestions and technical assistance. We are very grateful to the editor and reviewers for providing constructive comments for the improvement of our manuscript. We also thank for the fund supporting by the National Natural Science Foundation of China.
Authors' information

1. Institute of Bast Fiber Crops and Center of Southern Economic Crops, Chinese Academy of Agricultural Sciences, Changsha 410205, China. 2. The Open University of Hunan, Changsha 410004, China. 3. Faculty of Life Science and Technology, Central South University of Forestry and Technology, Changsha 410004, China. 4. Joint Laboratory of IBFC(China) and UOW(Poland) for Molecular Biology of Bast Fiber Crops, Changsha 410205, China.

References


### Tables

Table 1 Top twenty of down-regulated and up-regulated genes in pairwise comparisons of ABP30 vs. UNP30 and UNP 30 vs. UNP 50, respectively.
<table>
<thead>
<tr>
<th>Gene ID</th>
<th>log2Fold</th>
<th>Gene</th>
<th>Functional annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABP 30 vs UNP 30 (down-regulated)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lus10004281.g</td>
<td>-14.902</td>
<td>RHIE</td>
<td>Rhamnogalacturonate lyase</td>
</tr>
<tr>
<td>Lus10011565.g</td>
<td>-11.701</td>
<td>ACS8</td>
<td>1-aminocyclopropane-1-carboxylate synthase 8</td>
</tr>
<tr>
<td>Lus1002984.g</td>
<td>-11.696</td>
<td>FLA11</td>
<td>Fasciclin-like arabinogalactan protein 11</td>
</tr>
<tr>
<td>Lus10033672.g</td>
<td>-11.132</td>
<td>LRX4</td>
<td>Leucine-rich repeat extensin-like protein 4</td>
</tr>
<tr>
<td>Lus10020994.g</td>
<td>-10.567</td>
<td>PER10</td>
<td>Peroxidase 10</td>
</tr>
<tr>
<td>Lus10012808.g</td>
<td>-10.556</td>
<td>RCOM_0679870</td>
<td>CASP-like protein 1D1</td>
</tr>
<tr>
<td>Lus10015252.g</td>
<td>-10.363</td>
<td>SRG1</td>
<td>Protein SRG1</td>
</tr>
<tr>
<td>Lus10008932.g</td>
<td>-10.278</td>
<td>MLP328</td>
<td>MLP-like protein 328</td>
</tr>
<tr>
<td>Lus10036112.g</td>
<td>-10.063</td>
<td>FLA12</td>
<td>Fasciclin-like arabinogalactan protein 12</td>
</tr>
<tr>
<td>Lus10036113.g</td>
<td>-9.1933</td>
<td>FLA12</td>
<td>Fasciclin-like arabinogalactan protein 12</td>
</tr>
<tr>
<td>Lus10016080.g</td>
<td>-9.0561</td>
<td>At3g07570</td>
<td>Cytochrome b561 and DOMON domain-containing protein At3g07570</td>
</tr>
<tr>
<td>Lus10031326.g</td>
<td>-8.9001</td>
<td>MYB44</td>
<td>Transcription factor MYB44</td>
</tr>
<tr>
<td>Lus10035585.g</td>
<td>-8.8041</td>
<td>At5g03810</td>
<td>GDSL esterase/lipase At5g03810</td>
</tr>
<tr>
<td>Lus10029145.g</td>
<td>-8.7309</td>
<td>GOLS1</td>
<td>Galactinol synthase 1</td>
</tr>
<tr>
<td>Lus10022577.g</td>
<td>-8.6874</td>
<td>PIP2-1</td>
<td>Aquaporin PIP2-1 OS=Zea mays GN=PIP2-1 PE=1 SV=2//1.33266e-19</td>
</tr>
<tr>
<td>Lus10026200.g</td>
<td>-8.6063</td>
<td>NAC021</td>
<td>NAC domain-containing protein 21/22</td>
</tr>
<tr>
<td>Lus10005818.g</td>
<td>-8.597</td>
<td>ATL76</td>
<td>E3 ubiquitin-protein ligase ATL76</td>
</tr>
<tr>
<td>Lus10028638.g</td>
<td>-8.5858</td>
<td>MYB44</td>
<td>Transcription factor MYB44</td>
</tr>
<tr>
<td>Lus10009932.g</td>
<td>-8.4871</td>
<td>GSVIVT00023967001</td>
<td>Peroxidase 4</td>
</tr>
<tr>
<td>Lus10012663.g</td>
<td>-8.4764</td>
<td>NPF5.7</td>
<td>Protein NRT1/ PTR FAMILY 5.7</td>
</tr>
<tr>
<td><strong>UNP 30 vs UNP 50 (up-regulated)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lus10000527.g</td>
<td>Inf</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Lus10000773.g</td>
<td>Inf</td>
<td>At2g41970</td>
<td>Probable protein kinase At2g41970</td>
</tr>
<tr>
<td>Lus10000834.g</td>
<td>Inf</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Lus10000835.g</td>
<td>Inf</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Lus10000836.g</td>
<td>Inf</td>
<td>HR3</td>
<td>RPW8-like protein 3</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Inf</td>
<td>Gene Name</td>
<td>Description</td>
</tr>
<tr>
<td>------------------</td>
<td>-----</td>
<td>-------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Lus10000857.g</td>
<td>Inf</td>
<td>At1g77330</td>
<td>1-aminocyclopropane-1-carboxylate oxidase 5</td>
</tr>
<tr>
<td>Lus10002396.g</td>
<td>Inf</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Lus10002397.g</td>
<td>Inf</td>
<td>SBT1.6</td>
<td>Subtilisin-like protease SBT1.6</td>
</tr>
<tr>
<td>Lus10003148.g</td>
<td>Inf</td>
<td>CYP82C4</td>
<td>Cytochrome P450 82C4</td>
</tr>
<tr>
<td>Lus10003451.g</td>
<td>Inf</td>
<td>MLP423</td>
<td>MLP-like protein 423</td>
</tr>
<tr>
<td>Lus10004046.g</td>
<td>Inf</td>
<td>At4g08850</td>
<td>Probable LRR receptor-like serine/threonine-protein kinase At4g08850</td>
</tr>
<tr>
<td>Lus10004452.g</td>
<td>Inf</td>
<td>HD3A</td>
<td>Protein HEADING DATE 3A</td>
</tr>
<tr>
<td>Lus10005104.g</td>
<td>Inf</td>
<td>ATL40</td>
<td>RING-H2 finger protein ATL40</td>
</tr>
<tr>
<td>Lus10005178.g</td>
<td>Inf</td>
<td>FMO1</td>
<td>Probable flavin-containing monooxygenase 1</td>
</tr>
<tr>
<td>Lus10006692.g</td>
<td>Inf</td>
<td>ROMT</td>
<td>Trans-resveratrol di-O-methyltransferase</td>
</tr>
<tr>
<td>Lus10007073.g</td>
<td>Inf</td>
<td></td>
<td>Agglutinin-2</td>
</tr>
<tr>
<td>Lus10007890.g</td>
<td>Inf</td>
<td></td>
<td>Miraculin</td>
</tr>
<tr>
<td>Lus10007892.g</td>
<td>Inf</td>
<td></td>
<td>Miraculin</td>
</tr>
<tr>
<td>Lus10008019.g</td>
<td>Inf</td>
<td></td>
<td>Unknown</td>
</tr>
<tr>
<td>Lus10008083.g</td>
<td>Inf</td>
<td>BOR4</td>
<td>Boron transporter 4</td>
</tr>
</tbody>
</table>

“Inf” is abbreviation of infinity.

**Table 2** Top twenty of down-regulated genes in pairwise comparisons of ABP50 vs. UNP50 and UNP 30 vs. UNP 50.
<table>
<thead>
<tr>
<th>Gene ID</th>
<th>log2Fold</th>
<th>Gene</th>
<th>Functional annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABP50 vs UNP 50 (down-regulated)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lus10008930.g</td>
<td>-7.9134</td>
<td>MLP328</td>
<td>MLP-like protein 328</td>
</tr>
<tr>
<td>Lus10002953.g</td>
<td>-5.8702</td>
<td>ERF022</td>
<td>Ethylene-responsive transcription factor ERF022</td>
</tr>
<tr>
<td>Lus10033555.g</td>
<td>-5.3647</td>
<td>CYP735A1</td>
<td>Cytokinin hydroxylase</td>
</tr>
<tr>
<td>Lus10026527.g</td>
<td>-5.2991</td>
<td>NRT2.4</td>
<td>High affinity nitrate transporter 2.4</td>
</tr>
<tr>
<td>Lus10034505.g</td>
<td>-5.2718</td>
<td>CYP83B1</td>
<td>Cytochrome P450 83B1</td>
</tr>
<tr>
<td>Lus10006307.g</td>
<td>-4.9387</td>
<td>SBT1.7</td>
<td>Subtilisin-like protease SBT1.7</td>
</tr>
<tr>
<td>Lus10027412.g</td>
<td>-4.8852</td>
<td>DREB1D</td>
<td>Dehydration-responsive element-binding protein 1D</td>
</tr>
<tr>
<td>Lus10011872.g</td>
<td>-4.8673</td>
<td>SDI1</td>
<td>Protein SULFUR DEFICIENCY-INDUCED 1</td>
</tr>
<tr>
<td>Lus10015036.g</td>
<td>-4.7284</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Lus10032218.g</td>
<td>-4.708</td>
<td>UGT85A24</td>
<td>7-deoxyloganetin glucosyltransferase</td>
</tr>
<tr>
<td>Lus10006668.g</td>
<td>-4.5712</td>
<td>yoaA</td>
<td>Uncharacterized N-acetyltransferase YoaA</td>
</tr>
<tr>
<td>Lus10030554.g</td>
<td>-4.5654</td>
<td>At3g19850</td>
<td>BTB/POZ domain-containing protein At3g19850</td>
</tr>
<tr>
<td>Lus10021473.g</td>
<td>-4.3737</td>
<td>Tmem205</td>
<td>Transmembrane protein 205</td>
</tr>
<tr>
<td>Lus10004180.g</td>
<td>-4.3724</td>
<td>FAD2</td>
<td>Omega-6 fatty acid desaturase, endoplasmic reticulum</td>
</tr>
<tr>
<td>Lus10029838.g</td>
<td>-4.2832</td>
<td>LSH10</td>
<td>Protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 10</td>
</tr>
<tr>
<td>Lus10020700.g</td>
<td>-4.1491</td>
<td>LSH10</td>
<td>Protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 10</td>
</tr>
<tr>
<td>Lus10004181.g</td>
<td>-3.9629</td>
<td>FAD2</td>
<td>Omega-6 fatty acid desaturase, endoplasmic reticulum</td>
</tr>
<tr>
<td>Lus10015206.g</td>
<td>-3.961</td>
<td>PPC6-7</td>
<td>Probable protein phosphatase 2C 73</td>
</tr>
<tr>
<td>Lus10041692.g</td>
<td>-3.8414</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Lus10026948.g</td>
<td>-3.7975</td>
<td>HKT1</td>
<td>Sodium transporter HKT1</td>
</tr>
<tr>
<td><strong>UNP30 vs UNP 50 (down-regulated)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lus10010482.g</td>
<td>-9.9803</td>
<td>TPRP-F1</td>
<td>36.4 kDa proline-rich protein</td>
</tr>
<tr>
<td>Lus10000723.g</td>
<td>-9.1419</td>
<td>TMEM205</td>
<td>Transmembrane protein 205</td>
</tr>
<tr>
<td>Lus10015276.g</td>
<td>-7.1798</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Lus10008992.g</td>
<td>-7.0571</td>
<td>SAUR21</td>
<td>Auxin-responsive protein SAUR21</td>
</tr>
<tr>
<td>Lus10033722.g</td>
<td>-6.8499</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Lus10021473.g</td>
<td>-6.8354</td>
<td>Tmem205</td>
<td>Transmembrane protein 205</td>
</tr>
<tr>
<td>Lus10033038.g</td>
<td>-6.7895</td>
<td>BT1</td>
<td>BTB/POZ and TAZ domain-containing protein 1</td>
</tr>
<tr>
<td>Lus10002668.g</td>
<td>-6.7359</td>
<td>Anthranilate N-methyltransferase</td>
<td></td>
</tr>
<tr>
<td>Lus10007411.g</td>
<td>-6.7104</td>
<td>SYP71</td>
<td>Syntaxin-71</td>
</tr>
<tr>
<td>Lus10025321.g</td>
<td>-6.4556</td>
<td>TIC32</td>
<td>Short-chain dehydrogenase TIC 32, chloroplastic</td>
</tr>
<tr>
<td>Lus10010758.g</td>
<td>-6.4018</td>
<td>CYP82A1</td>
<td>Cytochrome P450 82A1 (Fragment)</td>
</tr>
<tr>
<td>Lus10035402.g</td>
<td>-6.3424</td>
<td></td>
<td>Nitrate reductase [NADH]</td>
</tr>
<tr>
<td>Lus10013956.g</td>
<td>-6.2354</td>
<td>At5g45950</td>
<td>GDSL esterase/lipase At5g45950</td>
</tr>
<tr>
<td>Lus10023515.g</td>
<td>-6.0237</td>
<td>PIP2-5</td>
<td>Probable aquaporin PIP2-5</td>
</tr>
<tr>
<td>Lus10034505.g</td>
<td>-5.9872</td>
<td>CYP83B1</td>
<td>Cytochrome P450 83B1</td>
</tr>
<tr>
<td>Lus10000868.g</td>
<td>-5.9356</td>
<td>CPP1</td>
<td>Protein CHAPERONE-LIKE PROTEIN OF POR1, chloroplastic</td>
</tr>
<tr>
<td>Lus10004886.g</td>
<td>-5.897</td>
<td>CET4</td>
<td>CEN-like protein 4</td>
</tr>
<tr>
<td>Lus10029838.g</td>
<td>-5.87</td>
<td>LSH10</td>
<td>Protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 10</td>
</tr>
<tr>
<td>Lus10041484.g</td>
<td>-5.7543</td>
<td>alx-1</td>
<td>Apoptosis-linked gene 2-interacting protein X 1</td>
</tr>
<tr>
<td>Lus10032218.g</td>
<td>-5.631</td>
<td>UGT85A24</td>
<td>7-deoxyloganetin glucosyltransferase</td>
</tr>
</tbody>
</table>

**Figures**
Figure 1

Scheme of sample collection. Segments of flax stem top region (ABP) and middle region (UNP) were isolated when plants were about 30cm and 50cm, respectively. SP: Snap point. Bar=1cm.
Figure 2

RNA-Seq analysis of sixteen stem samples and different expression genes across four types of stem samples. (A) Summary of RNA-Seq mapped to reference genes. (B) The correlation of RPKM values of all genes between the four replicates of ABP30, ABP50, UNP30, UNP50. (C) Number of up-regulated and down-regulated genes in the case of pairwise analysis of ABP30, ABP50, UNP30, UNP50 samples of flax stem. (D) Venn diagram showing distribution and number of specific, common and uniquely expressed genes in the case of a pairwise analysis of ABP30 vs. ABP50 (blue ellipse), ABP30 vs. UNP50 (green ellipse), ABP50 vs. UNP50 (yellow ellipse), UNP30 vs. UNP50 (purple ellipse).
Figure 3

Gene classification based on Gene Ontology (GO) for differentially expressed genes (DEGs) in four types of stem samples.
Figure 4

Kyoto encyclopedia of genes and genomes database (KEGG) functional annotations of differentially expressed genes (DEGs) for four types of stem samples.

(A) CesA

Glucan chain

Plasma membrane

Korrigan cellulose

Microtubules

Sucrose synthase

UDP-Glc

(B) Lus10007206g CESA6 Cellulose synthase activity
Lus10006224g CESA6 Cellulose synthase activity
Lus10005095g CSLG1 Cellulose synthase activity
Lus10005096g CSLG4 Cellulose synthase activity
Lus10007206g CESA6 Cellulose synthase activity
Lus10007198g MBP1 Cellulose biosynthetic process
Lus10007205g CSLG3 Cellulose biosynthetic process
Lus10004208g rhlE Cellulose biosynthetic process
Lus10016825g CSLG1 Cellulose metabolic process
Lus10004665g UPTG2 Cellulose metabolic process
Lus10004664g UAM1 Cellulose metabolic process
Lus10002414g glmM Cellulose binding
Lus10004670g COBL4 Cellulose microfibril organization
Lus10017663g COBL4 Cellulose microfibril organization
Lus10003197g COBL4 Cellulose microfibril organization
Lus10005113g COBL4 Cellulose microfibril organization
Lus10012162g KOR Cellulose microfibril organization
Lus10003083g CALS1?
Cellulose biosynthesis and expression profiles of genes involved in cellulose biosynthesis. (A) Plant’s cellulose biosynthesis: CesA complexes (CSC) use Uridine diphosphate-glucose (UDP-Glc) as the activated sugar donor for the β-1,4 glucan chain polymerization. UDP-Glc synthesized by sucrose synthase. Plasma membrane-bound forms of sucrose synthase are implicated in carbon supply for cellulose production. Korrigan cellulase (Kor) has been involved in monitoring of cellulose synthesis. Microtubules (MT) play role to regulate CesA proteins trafficking (Amor et al., 1995, Doblin et al., 2002, Endler and Persson, 2011). (B) Expression profiles of genes involved in cellulose biosynthesis.

Figure 6

Comparison of CESA and CSL genes expression values obtained by qRT-PCR and RNA-Seq analysis. (A) Quantitative RT-PCR validation of 10 differentially expressed CESA and CSL genes among ABP30, ABP50,
UNP30 and UNP50 revealed by digital gene expression (DGE) libraries of Linum usitatissimum L.. The values for the samples are presented as the fold change in gene expression, normalized to the CYC gene. RPKM: Gene expression levels were measured as reads per kilobase of exon model per million mapped reads. CESA: Cellulose synthase A; CSLG, CSLD and CSLE: Cellulose synthase-like protein. Vertical bars represent the mean ± SD (n=4). (B) Overall and normalized form of expression between qRT-PCR and FPKM for differentially expressed CESA and CSL genes among ABP30, ABP50, UNP30 and UNP50.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx